

Amendments to the claims are indicated in the attached "Marked Up Version of Amendments" (page i).

### REMARKS

Claims 1-62 remain in the application. No claim has yet been allowed.

The patent Examiner had indicated that claims 59-61 were patentable over the prior art and should be rewritten in independent form. With entry of the foregoing amendment to these claims, they should now be allowable.

Claims 62 now depends from claim 59 and thus is allowable for the same reason.

#### Claim 1 is not anticipated or obvious in view of Ahmad

Most of the remaining claims were rejected as being anticipated or obvious in view of U.S. Patent 6,263,507 issued to Ahmad et al. (which the Examiner calls the "Ahmed" reference).

Ahmad teaches a system for separating video or audio by first performing a coarse grained, then fine grained segmentation (Ahmad, col. 8 lines 22-26; col. 26 lines 62-67 and col. 27 lines 1-28). He does this based on, for example, word frequency similarities within segments.

In contrast, the present invention exploits a (manually crafted or machine learned) formal and declarative model of a news program to include named types or classes of segments. These segment classes may for example include broadcast initiation, anchor segment, (story) highlight segment, reporter segment, start of story, end of story, advertisement, broadcast termination, and so on. Probabilistic and temporal transitions among these segment types or classes in the model are based upon analysis of large scale collections (i.e., thousands of examples) of broadcast news programs. Because of this broad scale analysis, the present invention is to provide a more comprehensive and more specific set of segment cues than in Ahmad, but more importantly, it is able to increase segmentation performance by exploiting the model.

The invention is capable of modeling broadcast program types (e.g., CNN Prime News, CNN with Aaron Brown, Lehrer News Hour) with a declarative Finite State Machine (FSM) which is augmented with temporal intervals. This Time Enhanced Finite State Machine (T-FSM) is possibly in and of itself a more broadly applicable invention (See Figure 10, 11, and 12 in the patent application). The T-FSM is a recognition engine then takes in the many different types of multimedia "cues" (some of which were identified in Ahmad) but then compare their time sequence to recognize the structure of a specific instance of a news program type. This recognized structure (i.e., the T-FSM recognized for a specific news program on a particular evening) is then exploited using general methods (using techniques that apply to all instances of that type of program) to extract, summarize, and present news digitally and in a non-linear, directly searchable fashion.

Subject based segmentation of multimedia sources such as broadcast news video (i.e., the primary method in Ahmad) is in fact less effective (less precise in finding stories) than the invention, which provides a superior model based segmentation for a number of reasons including the short length of stories and errors in transcription (e.g., human errors from manual closed captions or machine errors from imperfect automated speech recognition transcription). Note that Ahmad et al. clearly disclose topic or subject oriented news stories (e.g., "world", "national", "local", "business", "sports", "living"), as detailed in column 15, lines 25-40.

In short, the T-FSM model is essential in two key ways:

- During news understanding it helps the system classify current events and anticipate subsequent ones which increases the performance (precision and recall) of story segmentation as documented in the above publication.

- During retrieval, recognized news structure is exploited in interesting subsequent ways. For example, cue frames are selected from the beginning of anchor segments but from the middle of reporter segments (segment types are not recognized in Ahmad as detailed below). Also, the most frequent named entities within stories are used to represent stories (NEs are pointed out but no invention for their extraction is disclosed in Ahmad)

(U) In summary, then, Ahmad does not teach one to form a model of a multimedia presentation based upon event cues, and then classifying that presentation based upon how well it matches such a model, as recited in claim 1 element (a).

And Ahmad does admittedly teach correlating words to classify the SUBJECT matter of a segment (e.g., news, sports, etc.), while also looking at "cues" to break a multimedia stream into segments. However, those two features do not amount to developing a model of expected events in a presentation such as a broadcast news program. And Ahmad certainly has no notion of all of matching detected events cues against such a model as recited in element (b) of claim 1.

Therefore, since claim 1 contains multiple features that are not suggested by Ahmad, claim 1 and all claims that depend from it are patentable.

Claims 14, 15, 16, 20, 21, 23, 38, 39, 41 – "Named Entities"

There is an important distinction between keywords and named entities that is missed in the Examiner's analysis. Amhad et al. provide neither description, nor detail, nor evidence that they use any natural language processing to extract named entities, other than a cursory example ("Jane Doe Anytown, USA", col. 23, lines 55-59). They are apparently processing keywords. However, keywords are text strings, typically occurring in a manually created pre-existing list which are often expanded using simple pattern matching and/or lexical (e.g., prefix and suffix processing) analysis techniques (e.g., so variants such as "unite", "unites" and "united" can be treated equivalently).

In contrast, named entities as claimed, are proper nouns such as people, organizations, locations, facilities, and so on. These are recognized by complex language understanding methods (e.g., rules, hidden Markov models) that reason at many levels including patterns (e.g., titles such as "Dr.", "Mr." or "Mrs."), part of speech (e.g., sequences of proper nouns might suggest a name), models of syntax (e.g., last names typically follow first) and models semantics (e.g., occupations are common last names as in "John Carpenter" or "Sally Blacksmith"; locations may be part of names originating from where a person was from as in "Jose de Seville", organization names may consist of complex subparts (e.g., "Department of Health/Motor Vehicles/Agriculture/ ...")). In this case the inventors have used machine learning techniques to train algorithms to detect named entities in transitions such as anchor to report, report to anchor, and so on. Using algorithms that can detect a class of named entity (e.g., a <person>, <organization>, or <location>) enables one to author or machine learn very

powerful patterns such as “I’m <person>” or “This is <person> reporting from <location>”). These patterns again were discovered by observing human annotators and are not obvious nor easily implemented. Because of the state of the art in language processing at the time, we do not believe Ahmad et al. had access to this level of sophisticated named entity extraction.

#### Claim 2 – Sentence Location for Summarization

(2) Ahmad et al. does, as the Examiner points out, detect segments that have similarity to prior segments including keyword frequency similarity. As described above, however, keywords are distinct from named entities, which are classes of linguistic expressions including names, organizations, etc. Moreover, Ahmad performs this word analysis to integrate similar story segments. In contrast, the present invention uses named entities to extract sentences from news segments, not to integrate story segments. Again, this is motivated by careful observation and analysis of human story annotators.

#### Claim 49 – Named Entities Summary

The example given by the Examiner from Ahmad (See “Erin: A Tropical Storm” in Fig 2-b) does not use or mention named entities.

#### Claim 50 – Sentence Summary

The example given by the Examiner from Ahmad (See “Erin: A Tropical Storm” in Fig 2-b) also does not use or suggest named entities to extract a sentence to serve as a summary.

#### Claim 53 – Named Entity Search

Because named entities are distinct from keywords, this claim is also distinct from prior art.

#### Claims 44, 45, 46 – Key Frames

While key frame extraction is not novel and disclosed by Ahmed and other prior art, what is novel here is the extraction of key frames using heuristics based on the type of segment. For example, as claimed, in the invention we perform the following for each type of segment:

- anchor segment – extract key frame from middle where story often is displayed or identified by graphics behind the anchor (claim #46)
- reporter segment – extract key frame from middle of segment where story content displayed as reporter is less significant (claim #45)
- advertisements/commercials – no key frame extract

The Examiner references Ahmad at col. 24 lines 37-41 which addresses cue phrases used for segmentation. This is not the same thing as the use of a state in a news model (in this case reporter or anchor booth segment) that then drives the selection of the keyframe from the middle of that type of segment.

In fact, Ahmad (not in the referenced section but rather in col. 17, lines 50-57) states that for television news “a video frame that occurs one tenth of the way through the video data representing the news story is selected.” Thus all new stories are treated equivalently by Ahmad, which would be problematic for overview segments or reporter segments where often the best content occurs in the middle of the segment (e.g., cameraman panning from the reporter in the street to the fire burning behind them).

#### Claim 17– Introductory news broadcast terms

Ahmad’s “update from” as referenced at col. 24 lines 37-41 is not an introductory term. This state in our news broadcast model includes terms such as “I’m”, “welcome” etc. as detailed on page 6 lines 17-20 of the specification.

#### Claims 19, 20, 21, 22, 23, 24, 25, 26, 27 - Discourse cues

Discourse cues indicating story segments are linguistic “patterns” that like named entities are machine learned from an annotated corpus of news segments, as opposed to simple lists of words or word patterns. Examples of states detected, include cues indicating:

- program start (e.g., “Good evening, I’m” <person-name>, “hello from” <location-name>) – See table 1 in patent application.
  - anchor to reporter handoffs (“We go now to” <person-name> “in” <location-name>, <person-name> “reports”) - See table 3 in patent application.
  - reporter to anchor handoffs (“thank you” <person-name>; “I’m” <person-name> “in” <location-name> “reporting for”) - See table 4 in patent application.
  - story termination (e.g., “coming up next”) - See table 5 in our patent application
- Ahmad has no such notions.

#### Claims 34, 35 – Discourse silence

While it is true that Ahmad discloses the use of silence to detect a segment, the present invention recognizes specific intervals from automated analysis of large scale collections of broadcast news data, namely discovering that silence of .7 seconds or greater is indicative of commercials (see page 17 columns 12-29 of the patent application).

#### Claims 30, 31 and 32-37 – Statistical distributions of Events and Cues

Statistical distributions of words over time (since story start) are highly indicative of story classes (e.g., weather, sports, etc.). Whereas Ahmad does perform statistical comparisons of story segments to adjacent ones in order to seek to collapse/integrate similar ones into one, the present invention performs statistical analysis across many programs (not stories) to discover likelihood models of key terms occurring at different times in a broadcast. This general distribution model then can be used as an additional source of evidence to classify a given segment. So while it is possible to have a weather story at the beginning of the news if a significant tornado or other dangerous weather event is occurring, it is much more likely that this will occur about 25+ minutes into a ½ hour news broadcast.

## Claims 55-57- Hierarchical presentation

Claim 55 was rejected also in view of Ahmad and further in combination with U.S. Patent No. 5,629,733 (Youman). As we have mentioned above, Ahmad does discuss the use of key phrases for segmentation. And as noted above, Ahmad also does detect segments that have similarity to prior segments, including key word frequency similarity.

However, Ahmad has no notion of extracting information representing a story segment, and certainly no notion of extracting a story segment from the source multimedia presentation where such information includes a summary representation of the story. All Ahmad suggests is to select a specific numbered frame as a key frame.

While Youman teaches displaying hierarchical levels of a program guide, he has no notion of hyperlinks permitting navigation among related story segments to navigate to a desired hierarchical level of representation as set forth in element d of claim 55.

In fact, Youman is just a type of electronic program schedule. He does not at all suggest techniques for representing an entire multimedia presentation (i.e., the entire television program) as a set of hierarchical story segments that can be navigated through a series of hierarchical hyperlinks.

As to claim 56, we note that Applicants' claim requires that the summary representation include named entities. For the reasons given above, Ahmad does not teach the use of named entities and therefore cannot be said to render this claim obvious.

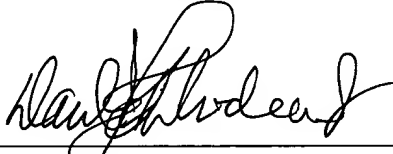
All of the claims should be allowed.

CONCLUSION

In view of the above amendments and remarks, it is believed that all claims are in conditions for allowance, and it is respectfully requested that the application be passed to issue. If after considering our remarks above, the Examiner still feels that the claims are to be rejected again, we would respectfully request the opportunity to meet with the Examiner in an interview or telephone conference call with the inventor.

Respectfully submitted,

HAMILTON, BROOK, SMITH & REYNOLDS, P.C.

By 

David J. Thibodeau, Jr.

Registration No.: 31,671

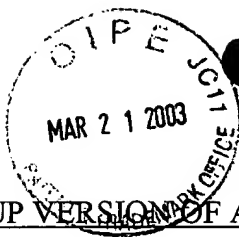
Telephone: (978) 341-0036

Facsimile: (978) 341-0136

Concord, MA 01742-9133

Dated: 3/17/2003



MARKED UP VERSION OF AMENDMENTS**RECEIVED**

MAR 28 2003

Technology Center 2600

Claim Amendments Under 37 C.F.R. § 1.121(c)(1)(ii)

59. (Amended) A method [as in Claim 58 additionally comprising the step of:] for automatically processing a representation of a multimedia presentation having multiple information streams contained therein, the method comprising the steps of:

- (a) selecting at least one contiguous portion of the multimedia presentation as a story segment;
- (b) extracting named entities from a text information stream corresponding to the story segment;
- (c) using extracted named entities as search criteria to select from among a plurality of story segments; and
- (d) in response to a search query, presenting a list of named entities and their corresponding number of occurrences in story segments over a selected time period.

60. (Amended) A method as in Claim [58] 59 additionally comprising the step of:  
[(d)] (e) in response to a search query, presenting a graph of named entities and their corresponding frequency of occurrences in story segments over a selected time period.

61. (Amended) A method as in Claim 60 additionally comprising the step of:  
[(e)] (f) in response to selection of a point on the graph of named entities, presenting the user with an overview story segments containing the selected named entity.

62. (Amended) A method as in Claim [58] 59 additionally comprising the step of:  
[(d)] (g) in response to a search query for a story segments of a selected type, presenting a thumbnail view comprising key frames from multiple story segments of the selected type.